

## Creating alternative access layers to the DOBES archive from existing metadata structure

Paul Trilsbeek

*Max Planck Institute for  
Psycholinguistics*

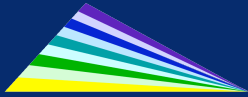
Gabriele Müller

*University of Münster*

Julia Colleen Miller

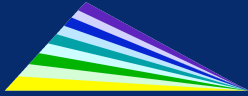
*University of Washington*

---



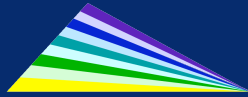
# Structure of the Talk

- Intro DoBeS Archive Structure
- Search/Browse Options
- Potential Problems for users
- Beaver Community Portal
- The Technology
- Further Entry Points into the Archive



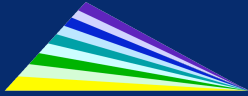
# DOBES Documentation Programme

- Endangered Languages documentation program funded by the German VW foundation
- Started in 1999 with a few pilot projects, until now 50 large documentation projects have been granted
- Central archive for all projects at Max Planck Institute in Nijmegen, about 45 Terabytes of data



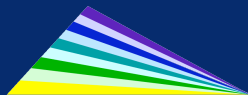
# DOBES Documentation Programme



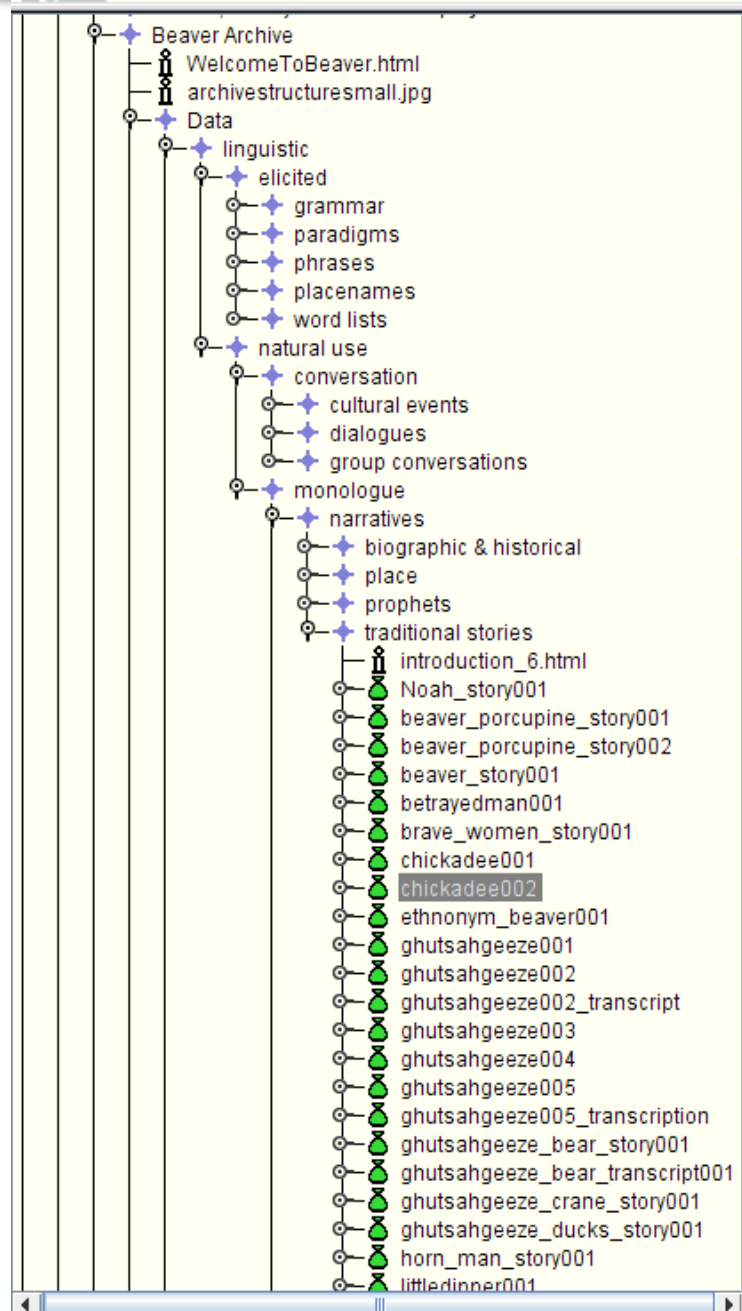


## DoBeS Archive: "Traditional Access"

- browsing by clicking on the nodes
- metadata search
- annotation content search



# DoBeS Archive



IMDI

ISLE Metadata Initiative

## Session

**Name** chickadee002  
**Title** story about chickadee  
**Date** 2006-02-15

## Description

story about chickadee

[Location](#)

[Project](#) Beaver

[Content](#)

## Actors

[Actor](#) Gabriele Mueller

[Actor](#) 891

[Actor](#) 908

[MediaFile](#)

[MediaFile](#)

[MediaFile](#)

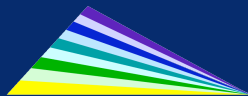
[MediaFile](#)

[WrittenResource](#)

[Source](#)

[Anonyms](#)

## References



# DoBeS Archive

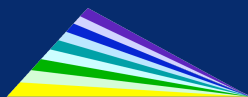
IMDI-Browser

[about](#) user: elekabele@mp

- Beaver Archive
  - WelcomeToBeaver.html
  - archivestructuresmall.jpg
  - Commentary
  - Community Materials and Homepages
  - Data
    - linguistic
      - Workshop 2008
      - elicited
      - natural use
        - conversation
        - monologue
        - narratives
          - biographic & historical
          - misc
          - place
          - prophets
          - traditional stories
            - introduction\_6.html
            - Noah\_story001
            - beaver\_porcupine\_story001
            - beaver\_porcupine\_story002
            - beaver\_story001
            - betrayedman001
            - brave\_women\_story001
            - chickadee001
            - chickadee002
            - ethnonym\_beaver001
            - ghutsahgeeze001
            - ghutsahgeeze002
              - ghutsahgeeze002.wav
              - ghutsahgeeze002\_VDP.mp4
              - ghutsahgeeze002\_VDP.mpeg
              - ghutsahgeeze002\_VDP.mpg
              - ghutsahgeeze002\_VDP.wav
              - ghutsahgeeze002\_transcript
            - ghutsahgeeze003
            - ghutsahgeeze004
            - ghutsahgeeze005
            - ghutsahgeeze005\_transcription
            - ghutsahgeeze\_bear\_story001
            - ghutsahgeeze\_bear\_transcript001
            - ghutsahgeeze\_crane\_story001
            - ghutsahgeeze\_ducks\_story001
            - horn\_man\_story001
            - littledipper001
            - littledinner002







# DoBeS Archive

## Metadata search

IMDI-Browser about user: anonymous

- ◆ Bakairi, Katxuyana and Mawe project
- ◆ Beaver Archive
  - ◆ WelcomeToBeaver.html
  - ◆ archivestructuresmall.jpg
- ◆ Commentary
- ◆ Community Materials and Homepages
- ◆ Data
  - ◆ linguistic
  - ◆ Workshop 2008
    - ◆ elicited
    - ◆ natural use
  - ◆ nonlinguistic
- ◆ Studies
- ◆ Cashinahua project
- ◆ Chaco languages
- ◆ Chintang and Puma Documentation Project
- ◆ Chipaya project
- ◆ Chipaya project II
- ◆ Chontal
- ◆ ECLING project
- ◆ Enets and Forest Nenets
- ◆ Gente del Centro - People of the Center
- ◆ Gorani
- ◆ Hocank team
- ◆ Iwaidja team
- ◆ Jaminjung and Eastern Ngumpin
- ◆ Kola Saami Documentation Project
- ◆ Kuikuro Team
- ◆ Lacandon Cultural Heritage
- ◆ Marquesan team
- ◆ Movima
- ◆ Salar Monguor Project
- ◆ Saliba/Logea
- ◆ Savosavo
- ◆ Semang

### Metadata search

☒ Show occurrences

within 1 selected corpus: linguistic [670 sessions]

Key word search **Standard search** Advanced search

Search in all fields:

Search

Save

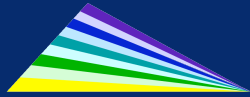
Print

Content search


13 matches within 670 selected sessions found. Search complete

- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/elicited/word lists/animals/moose\_bodyparts001
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/elicited/word lists/bodyparts003
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/narratives/biographic & historical/moose\_hunt001
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/narratives/place/snare\_hill\_sa-001
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/narratives/place/snare\_hill\_scenery-001
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/narratives/place/snare\_hill\_ta-001
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/narratives/traditional stories/moose\_story001
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/berry\_picking001
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/cut\_meat001
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/cut\_meat002
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/make\_flesher001
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/moose\_hide002
- /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/moose\_hide003





## Annotation content search

 ANNEX SEARCH [help](#)

user: [elekabele@mpi.nl](#) [logout](#)

Substring Search | Single Layer Search | Multiple Layer Search

Domain: Data

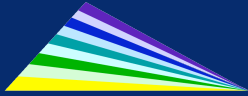
☒ EAF 80 ☒ Text 100

#hits : 220  
#annotations with a hit : 201  
#annotations investigated : 34123

>

hit 1 - 26 of 220

Our stuff is sitting back there" (said the wolf) "go and get it!" and that then/iguess??? a rib, a **moose** rib that's how they took him out (grabbed him with the rib and pulled him out) his is living now  
half grizzly xádaa **moose** tsits duck tlij  
strawberries xádaa jidze **moose**berries jít'qóq leaf jít'wáq  
'I hunt for **moose**' ghwu ghustl'u 'I'll  
'I shot a **moose**' aja kweì 'spear'  
xádaa tsán \tr **moose** meat \ref \t  
ijzaa September lit **moose** run month yaş  
'He killed a **moose**' xádaa ghazéxi 'They  
'They killed a **moose**' şj xádaa zeexi  
'I killed a **moose**' góq xádaa zíxi  
'You killed a **moose**' ukedúdye xádaa zásíghi  
'We killed the **moose**' ukedídye xádaa záàxi  
'Y'all killed the **moose**' Paradigm of "to  
I killed a **moose**' dyi dzene' xádaa  
I killed a **moose**'  
'puppy' dyééyuę 'bull **moose**' dyééyuę aadi 'I  
hear the bull **moose**' xádaa adzé' aadi  
hear the cow **moose**' chíye aadi 'I  
'I hear the **moose** calf' madzih 'caribou'  
horns' xádaa dvée' **moose** horns' advée' 'horns'

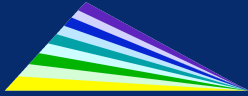


# DoBeS Archive: Potential Problems

- only in English
- very intricate metadata structure is only useful for specialist users
- not possible to know what is there at a glance
- too many clicks to find something
- community members will be easily discouraged to explore corpus

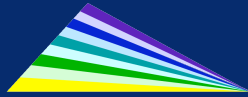
⇒ the data in the archive should be available to community

How to make access as easy as possible?

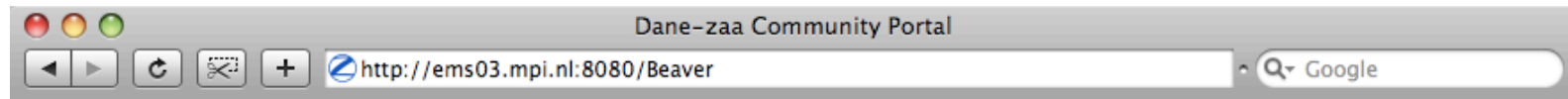


# The Idea: Community Portals

- provide a webinterface
- keep interface simple and attractive
- with focus on culturally important topics
- choose data in the archive that is relevant
- sort thematically
- use local language
- keep maintenance low, once the portal is set up
- new relevant resources should automatically appear



# Beaver Community Portal

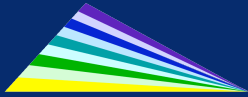


## Dane-zaa Community Portal

This page is created for the members of the Dane-zaa community to facilitate the use of the archive collected by the DoBeS team together with the elders.



Stories	Learn about ...	Materials	The Archive
Personal history	Drum	Movies	Searching
Traditional stories	Food and cooking	Clickables	Navigating
Animals	Handgames	Dictionary	Download
Place	Horses	Phrasebook	Tours
	Moccassins	Calendar	Google Earth
	Moosehide	Alphabet	Studies
	Preparing meat	Posters	



# Beaver Community Portal

Drum — Dane-zaa Community Portal


◀ ▶ ↺ ⌂ +


http://ems03.mpi.nl:8080/Beaver/learnabout/drum


🔍 Google


## Dane-zaa Community Portal

This page is created for the members of the Dane-zaa community to facilitate the use of the archive collected by the DoBeS team together with the elders.


Stories


Learn about...


Materials


The Archive

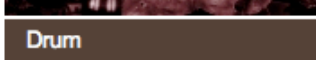
[Home](#) > [Learn about...](#) > [Drum](#)


Learn about...

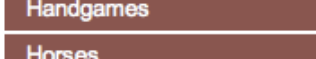
Drum


Food and Cooking

Handgames

Horses

Moccassins


Moosehide


Preparing meat


## Drum


Listen to drumming, learn about the different parts of a drum.

paradigm\_drum001 - paradigm drum

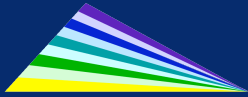
 paradigm\_drum001.m4a

 paradigm\_drum001.wav

 paradigm\_drum001.txt

 paradigm\_drum001.pdf

Go to "<http://ems03.mpi.nl:8080/Beaver/learnabout/>"




# Beaver Community Portal

Drum — Dane-zaa Community Portal

http://ems03.mpi.nl:8080/Beaver/learnabout/drum

Home > Learn about... > Drum



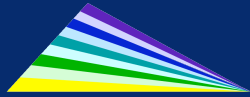
drum001.mpg

CLOSE X

drum002.wav

Loading "http://ems03.mpi.nl:8080/Beaver/learnabout/drum", completed 27 of 30 items (2 errors)






# General DOBES Portal

DOBES portal  
Documentation of Endangered Languages

Language Portal

This portal gives you access to some of the resources in the archive of the DOBES endangered languages programme. Choose a region below or a language or genre on the right.



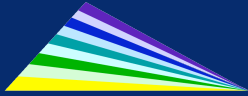
DOBES Languages »

- #Akhoe Hai /om
- Awetí
- Bakaíri
- Batek
- Baure
- Beaver
- Bora
- Cashinahua
- Chaco Languages
- Chintang
- Chipaya
- Chontal
- Gorani
- Hocank
- Iwaidja
- Jaminjung
- Katxuyana
- Kola-Sámi
- Kuikuro
- Lacandon
- Logea
- Marquesan
- Mawé
- Monguor
- Movima
- Nenets

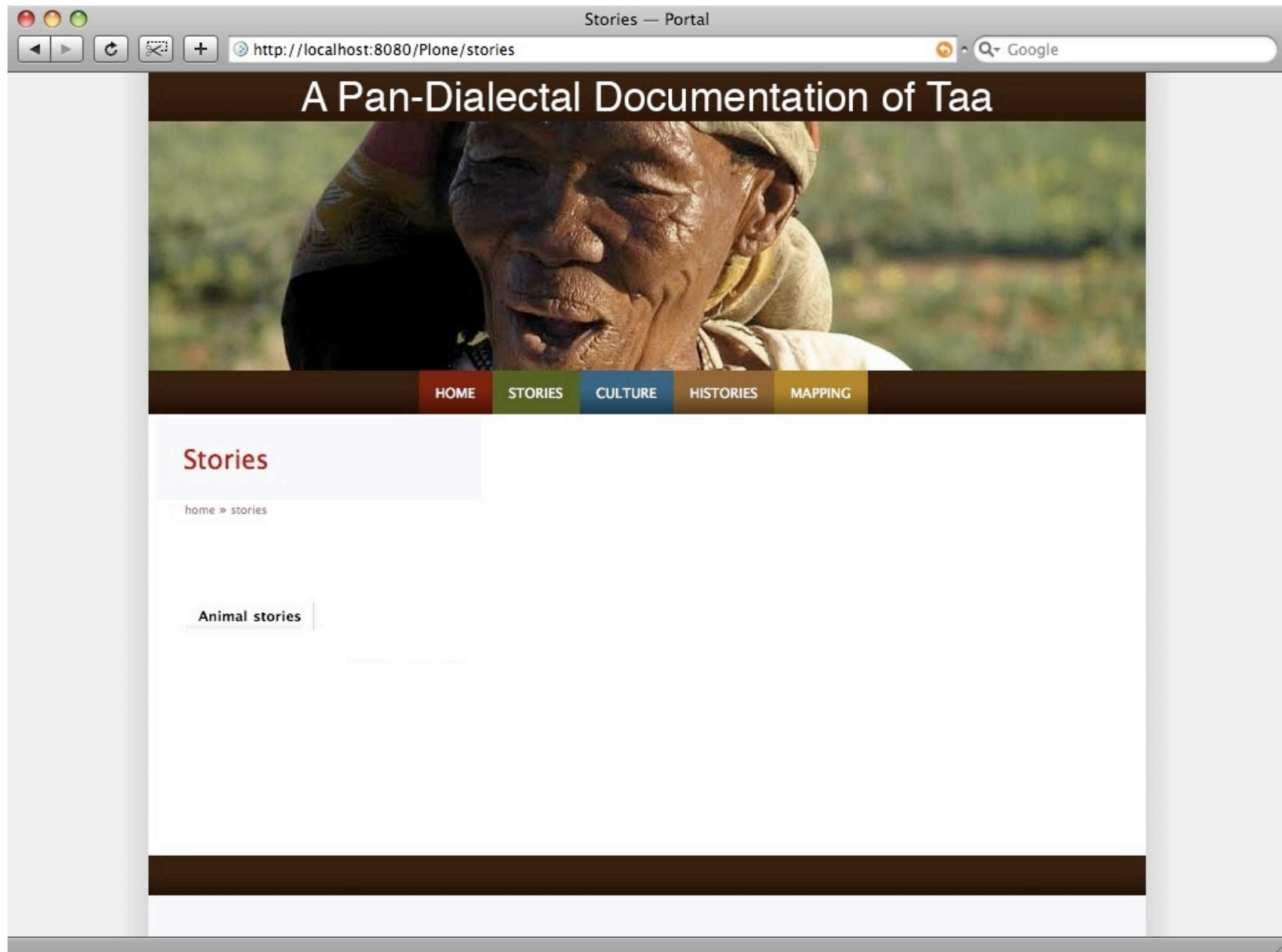
English

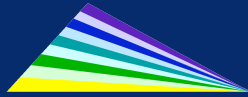
About »

This is a [WordPress](#) site. Design based on the [Fjords01!](#) theme.



# Taa portal






# OLCAP Portal

Welcome to OLCAP. — OLCAP


<http://olcap.aiatsis.gov.au/main> Google

 **OLCAP**

## Welcome to OLCAP.

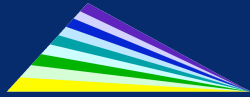
**Click on the image to the right to go to the portals.**

**Click here to go to the main AIATSIS website.**



Map copyright Geoscience Australia 2008.

Go to "<http://olcap.aiatsis.gov.au/vrd>"




# OLCAP Portal

People — OLCAP — Victoria River District

http://olcap.aiatsis.gov.au/vrd/gurindji/peoplesearch

Google

 **OLCAP**

Victoria River District

Bilinarra

Gurindji

**People**

Kinds

Titles

Jaminjung

Ngaliwurru

Ngarinyman

Wardaman

**People**

Bernard Peanut Japalyi

Biddy Wave Hill Nangala

Dandy Danbayarri Jimija

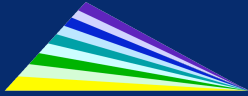
Harry George Japalyi

Helen Morris Namija

Lauren Campbell

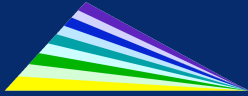
Ronnie Wave Hill Jangala

Violet Donald Nanaku



# Community Portals

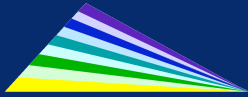
- Technology behind the portals needs to be generic enough to allow for a large variety of portals to be built with the same framework
- Design, layout and selection of search criteria need to be provided by the researchers in collaboration with the speech community
- The archiving team in Nijmegen technically implements the portal and hosts it (if desired)



# Metadata Search Web Service

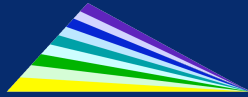
- REST (REpresentational State Transfer) Web Service
- Query the metadata database via a URL containing a number of parameters
- Search result comes back as XML, which can be further processed inside a dynamic website
- New data that is added to the repository and matches the search criteria automatically shows up in the portal





Example:

[http://corpus1.mpi.nl/ds/imdi\\_search/servlet?  
request=getMatches&num=50&query=Session.N  
ame="\\*drum\\*&type=advanced&nodeid=MPI790  
25%23&includeTitle=true&includeUrl=true&incl  
udeResources=true&returnType=xml](http://corpus1.mpi.nl/ds/imdi_search/servlet?request=getMatches&num=50&query=Session.Name=)

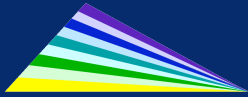


# Metadata Search Web Service

```
Mozilla Firefox
http://corpus1.mpi.nl/ds/imdi_search/servlet?request=getMatches&num=5
Google

- <Arguments>
  <Argument name="nodeid" Value="MPI79025#"/>
  <Argument name="includeTitle" Value="true"/>
  <Argument name="request" Value="getMatches"/>
  <Argument name="num" Value="50"/>
  <Argument name="includeUrl" Value="true"/>
  <Argument name="returnType" Value="xml"/>
  <Argument name="type" Value="advanced"/>
  <Argument name="query" Value="Session.Name=*drum*"/>
  <Argument name="includeResources" Value="true"/>
</Arguments>
<RequestProcessingTime>0.231</RequestProcessingTime>
</OperationRequest>
- <Result>
  <SessionId>F0B506B5BD1F6FEA6DF09C42B91DCCB2</SessionId>
  <SelectedSessionCount>688</SelectedSessionCount>
  <MatchCount>3</MatchCount>
  - <Match>
    <Id>MPI300270#</Id>
    <Name>paradigm_drum001</Name>
    <Title>paradigm of "to drum"</Title>
    - <TreePath>
      /IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/elicited/verbs/paradigms/paradigm_drum001
    </TreePath>
    - <URL>
      http://corpus1.mpi.nl/qfs1/media-archive/dobes_data/Beaver/Data/linguistic/elicited/verbs/paradigms/Metadata/paradigm_drum001.imdi
    </URL>
    - <Resource>
      <Id>MPI554715#</Id>
      <Name>paradigm_drum001.m4a</Name>
      <Format>audio/mp4</Format>
      - <URL>
        http://corpus1.mpi.nl/qfs1/media-archive/dobes_data/Beaver/Data/linguistic/elicited/verbs/paradigms/Media/paradigm_drum001.m4a
      </URL>
    </Resource>
  </Match>
</Result>
```

Done



# Beaver Community Portal

Drum — Dane-zaa Community Portal

[http://ems03.mpi.nl:8080/Beaver/learnabout/drum](#) [Google](#)

## Dane-zaa Community Portal

This page is created for the members of the Dane-zaa community to facilitate the use of the archive collected by the DoBeS team together with the elders.

Stories

Learn about...

Materials

The Archive

[Home](#) > [Learn about...](#) > [Drum](#)

### Learn about...

Drum

Food and Cooking

Handgames

Horses

Moccassins

Moosehide

Preparing meat

## Drum

Listen to drumming, learn about the different parts of a drum.

paradigm\_drum001 - paradigm drum

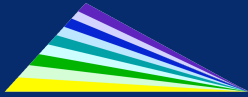
paradigm\_drum001.m4a

paradigm\_drum001.wav

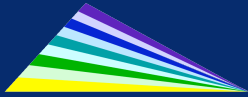
paradigm\_drum001.txt

paradigm\_drum001.pdf

[Go to "http://ems03.mpi.nl:8080/Beaver/learnabout"](#)



- Standard, open source Web Content Management System
- Written in Python
- Large number of add-on modules
- Built-in support for XML processing
- Written a custom content type for Plone: IMDI-query. Sends a query to the LAT server, result comes back as XML and is parsed into a nicely formatted list of resources



# New ANNEX

MPI – ANNEX Interface

[http://lux17.mpi.nl/ds/annex/protected/interface.jsp?jsessionid=2652E0914027FBB56D56296147E346BE](#) Google

**ANNEX** manual ? settings user: pautri@mpi.nl logout

**Text**  


---

**Grid**  


---

**Subtitle**  


---

**Waveform**  



---

**Timeline**  


---

**Combined**

Video display min



Media information min

Elapsed time: 00:00:45:658

Selected chunk:

Begin time: 00:00:42:000

End time: 00:00:43:000

Text: पर दुई नम्बर बाडमा त्यो गाउँ।

Player: Quicktime

Play selection  
Clear selection

Tier 1:  
nep@SR ▼  
Tier 2:  
eng@SR ▼  
Tier 3:  
none ▼  
Tier 4:

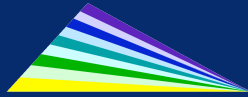
Subtitles

Tier 1: टेलीटोल, बालङ्गिटोलदेखि यता आउँदा खेर र्छि

Tier 2: Coming from the Waling settlement to this side to the Tele settlement.

Tier 3: [No source tier selected]

Tier 4: [No source tier selected]



# New ANNEX

MPI - ANNEX Interface

http://lux17.mpi.nl/ds/annex/protected/interface.jsp?sessionId=2652E0914027FBB56D56296147E346BE

Google

ANNEX

manual

?

settings

user: pautri@mpi.nl

logout

Text

Grid

Subtitle

Waveform

Timeline

Combined

Player: Quicktime

Play selection

Clear selection

Tier 1:

nep@SR

Tier 2:


none

Tier 3:

none

Font size:

Video display



⏮

⏪

⏩

⏭

Media information

Elapsed time: 00:00:43:000

Selected chunk:

Begin time: 00:00:42:000

End time: 00:00:43:000

Text: पर दुई नम्बर बाडमा त्यो गाउँ।

Tier 1

टूली आमा म आज पनसोध्छु । तपाईंहरु पर बस्नुहुन्थ्यो । पर मुलगाउँ बस्नुहुन्थ्यो। टेलेटोल मा बस्नुहुन्थ्यो । पर घेरै घरहरु थिए। यता त थिएनन् नति ! यो ! यता आउनु भयो । त्यहाँ पर त्य त्तबिला यहाँ अँ । यतातरि यहाँ । अँ । अँ । अँ । पर दुई नम्बर बाडमा त्यो गाउँ । टेलेटोल, बालडिडोल देखियता आउँदाखेरि यहाँ के के फलथ्यो ? यो ठाउँमा ? त्यसबेला यहाँ के थियो ? तरुलको लहरा गठि। भ्याक्कुर। अँ । अहलि पनयितातरि बन तरुल नसिक्छन् । बन तरुल खन्छौं । नसिक्छन् । त्यसो ह भने, पहलि पर छाडेर यता आईहालन् भएकै ठीक भयो । पर । घरहरु मात्रै । सुँगुरको लुँदीको थ पुरो। घेरै नै फलथ्यो होला । र त्यो मल, त्यो सुँगुरको लुँदी यता पनओसारन् भयो ? त्यसपछि

Shrink

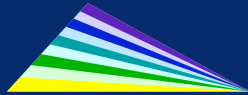
Tier 2

Expand

Tier 3

Expand





# New ANNEX

MPI – ANNEX Interface

http://lux17.mpi.nl/ds/annex/protected/interface.jsp?sessionId=2652E0914027FBB56D56296147E346BE

Google

ANNEX

manual

?

settings

user: pautri@mpi.nl

logout

Text

Grid

Subtitle

Waveform

Timeline

Combined

Player: Quicktime

Play selection

Clear selection

|<

>|

<<

>>

<


>

+

-

☐ Play screen by screen

Video display



⏮

⏪

⏩

⏭

Media information

Elapsed time: 00:00:09:000

Selected chunk:

Begin time: -

End time: -

Text: -

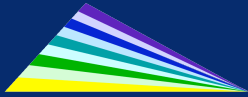
Mini Data Frame

Tier: none

Font size: 14

Timeline

eng@SR	You were living over there. you were living at mulgaun o
nep@SR	तपाईंहरू पर बस्नुहुन्थ्यो। पर मुलगाउँ बस्नुहुन्थ्यौ।
gram@SR	
cxt@SR	



# CuPED



...nochmol waut vetahle von, um, waut ekj aus Kjind erläwt hab.  
*...tell something again about, um, what I experienced as a child.*

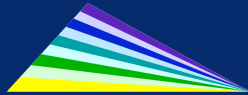
Daut wea, uh, em Somma emma sea woam.  
*It was, uh, always very warm in the summertime.*

Un, uh, uh, dee,  
*And, uh, uh, the,*

wi haude vâl Oft.  
*we had lots of fruit.*

Miene Taunte haud sea vâl Oft jedrieegt.  
*My aunt had dried a lot of fruit.*

Dee deed emma vâl Oft driejeje,  
*She always dried a lot of fruit,*



# Google Earth

Google Earth

File Edit View Tools Add Help

Search

Fly To Find Businesses Directions

Fly to e.g., Hotels near JFK

Places Add Content

My Places

- Beaver Placenames-FINAL.k...
- DoBeS\_Beaver-FINAL.kmz
- DoBeS\_Beaver
  - Introduction: Dane-
  - Doig River
  - Eleske
  - Snare Hill

Layers

- Primary Database
- Geographic Web
- Roads
- 3D Buildings
- Street View
- Borders and Labels
- Traffic
- Weather
- Gallery
- Ocean
- Global Awareness
- Places of Interest
- More
- Terrain

Snare Hill

Madáts'atl'oje "Snares Set Around It"

Sammy Acko shares a story about the importance of Madáts'atl'oje to the Dane-zaa people. The story is told in the Dane-zaa language, with English subtitles.

Introduction: Dane-zaa stories of place

Snare Hill

Alédzé

Wolf Creek

Berry River Reserve

Petersen's Crossing

Sweeney Creek

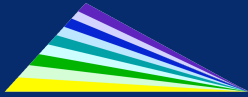
ENTER HERE Enter the archive!

Listen to the entire story!

DOBES Visit our homepage!

Directions: [To here](#) - [From here](#)

Google



## Conclusions

- Different user groups have different requirements regarding interfaces
- The same archival material can be presented in various ways, especially when making use of XML based formats
- Web Services make it easy to use archived material in different online contexts in a programmatic way